

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2003, 10:19:03 ; Search time 658.074 Seconds

(without alignments)  
11025.472 Million cell updates/sec

Title: US-09-922-895-4

Perfect score: 448  
Sequence: 1 TAGTCAGATGCAGAAAT.....CCCTGCTTAATGAAAGCTT 448

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estcin:\*  
4: em\_estmu:\*  
5: em\_estrov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| c 1        | 387.4 | 86.5        | 747    | 12    | BG220110 RST39882  |
| 2          | 76.4  | 17.1        | 655    | 10    | BB223728 BB223728  |
| 3          | 65.6  | 14.6        | 274    | 12    | BF553209 UI-R-C2-n |
| 4          | 49.4  | 11.0        | 320    | 10    | BB225464 BB225464  |
| 5          | 47.4  | 10.6        | 1203   | 17    | CNS015MU           |
| c 6        | 47.2  | 10.5        | 1027   | 17    | CNS02750           |

|      |      |      |      |    |           |                     |
|------|------|------|------|----|-----------|---------------------|
| c 7  | 46.4 | 10.4 | 1101 | 17 | CNS016LI  | AL106896 Drosophila |
| 8    | 46.2 | 10.3 | 914  | 17 | CNS002JY  | AL097768 Drosophila |
| 9    | 46   | 10.3 | 1101 | 17 | CNS017ZQ  | AL108704 Drosophila |
| 10   | 45.8 | 10.2 | 975  | 17 | CNS0102Y  | AL098452 Drosophila |
| 11   | 45.2 | 10.1 | 583  | 12 | BG603853  | BG603853 EST052943  |
| 12   | 44.6 | 10.0 | 874  | 17 | A2541515  | A2541515 ENTP060TF  |
| 13   | 44.6 | 10.0 | 1027 | 17 | CNS02750  | AL121273 Tetradon   |
| 14   | 44.4 | 9.9  | 1205 | 17 | CNS0165A  | AL106312 Drosophila |
| c 15 | 44.2 | 9.9  | 829  | 17 | BH501921  | BH501921 BOHJY48TR  |
| c 16 | 44.2 | 9.9  | 987  | 17 | CNS014PO  | AL104456 Drosophila |
| c 17 | 44.2 | 9.9  | 1101 | 17 | CNS0106X  | AL098595 Drosophila |
| c 18 | 43.6 | 9.7  | 699  | 17 | AQ323327  | AQ323327 Pct111-10  |
| 19   | 43.4 | 9.7  | 772  | 17 | CNS03AEN  | AL235112 Tetradon   |
| c 20 | 43.4 | 9.7  | 1086 | 17 | CNS00YXK  | AL096692 Drosophila |
| c 21 | 43.2 | 9.6  | 452  | 13 | BM273820  | BM273820 PEST0aa6   |
| c 22 | 43.2 | 9.6  | 1201 | 17 | CNS016FT  | AL106691 Drosophila |
| 23   | 43   | 9.6  | 546  | 10 | BB224066  | BB224066            |
| c 24 | 43   | 9.6  | 1101 | 17 | CNS000D1  | AL065414 Drosophila |
| 25   | 43   | 9.6  | 1101 | 17 | CNS00EO7  | AL069440 Drosophila |
| 26   | 42.8 | 9.6  | 540  | 13 | BJ077706  | BJ077706 Drosophila |
| c 27 | 42.8 | 9.6  | 1101 | 17 | CNS00LRT2 | AL078714 Drosophila |
| 28   | 42.6 | 9.5  | 964  | 17 | CNS007ER8 | AL441457 T7 end of  |
| 29   | 42.4 | 9.5  | 471  | 13 | BJ413351  | BJ413351            |
| 30   | 42.4 | 9.5  | 483  | 13 | BJ412970  | BJ412970            |
| 31   | 42.4 | 9.5  | 625  | 13 | BJ417045  | BJ417045            |
| 32   | 42.4 | 9.5  | 634  | 13 | BJ416550  | BJ416550            |
| 33   | 42.4 | 9.5  | 669  | 13 | BJ413381  | BJ413381            |
| c 34 | 42.2 | 9.4  | 567  | 13 | BJ410890  | BJ410890            |
| 35   | 42.2 | 9.4  | 1101 | 17 | CNS0004E  | AL067752 Drosophila |
| c 36 | 42   | 9.4  | 369  | 10 | AW411933  | AW411933 uo29g08.x  |
| c 37 | 42   | 9.4  | 458  | 17 | AQ408877  | AQ408877 HS_5067_A  |
| 38   | 42   | 9.4  | 499  | 13 | BJ324359  | BJ324359            |
| 39   | 42   | 9.4  | 556  | 13 | BJ388900  | BJ388900            |
| 40   | 42   | 9.4  | 656  | 13 | BJ387997  | BJ387997            |
| c 41 | 42   | 9.4  | 956  | 17 | CNS000KX2 | AL078376 Drosophila |
| c 42 | 41.8 | 9.3  | 484  | 17 | AQ866715  | AQ866715 nbeu0028L  |
| c 43 | 41.8 | 9.3  | 880  | 17 | BH152122  | BH152122 ENT0005TF  |
| c 44 | 41.8 | 9.3  | 948  | 17 | CNS0159T  | AL105179 Drosophila |
| c 45 | 41.8 | 9.3  | 951  | 17 | A2681741  | A2681741 ENTIP07TF  |

#### ALIGNMENTS

|            |  |             |      |        |                 |
|------------|--|-------------|------|--------|-----------------|
| RESULT 1   | BG220110/c   | 747 bp      | mRNA | linear | EST 21-APR-2001 |
| LOCUS      | BG220110   |             |      |        |                 |
| DEFINITION | RST39882 Athersys RAGE Library Homo sapiens CDNA, mRNA sequence.   |             |      |        |                 |
| ACCESSION  | BG220110   |             |      |        |                 |
| VERSION    | BG220110.1   | GI:13746131 |      |        |                 |
| KEYWORDS   | EST.   |             |      |        |                 |
| SOURCE     | human.   |             |      |        |                 |
| ORGANISM   | Homo sapiens   |             |      |        |                 |
| REFERENCE  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.   |             |      |        |                 |
| AUTHORS    | 1 (bases 1 to 747)<br>Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R., Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J., Denner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith,E., Veloso,N., Kika,A., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J. and Ducar,M. |             |      |        |                 |
| TITLE      | Creation of genome-wide protein expression libraries using random activation of gene expression  |             |      |        |                 |
| JOURNAL    | Nat. Biotechnol. 19 (5), 440-445 (2001)  |             |      |        |                 |
| MEDLINE    | 21227151   |             |      |        |                 |
| COMMENT    | Contact: Scott J. Cain<br>Athersys, Inc.<br>3201 Carnegie Ave, Cleveland, OH 44115, USA<br>Tel: 216 431 9900<br>Fax: 216 361 9596<br>Email: scain@atersys.com<br>High quality sequence stop: 523.<br>Location/Qualifiers   |             |      |        |                 |

FEATURES

|                           |   |  |
|---------------------------|---|--|
|                           | source  | 1..747<br>/organism="Homo sapiens"<br>/db_xref="taxon:9606"<br>/clone_lib="Altersys RAGE Library"<br>/cell_line="HT1080"<br>/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances." |
| BASE COUNT                | 192 a 142 c 183 g 230 t   |  |
| ORIGIN                    |   |  |
| Query Match               | 86.5% Score 387.4 DB 12 Length 747;   |  |
| Best Local Similarity     | 98.5% Pred. No. 7.5e-87;  |  |
| Matches 391; Conservative | 0; Mismatches 6; Indels 0; Gaps 0;  |  |
| OY                        | 1 TAGGTCAGATGCAGAAATTGGCTTAAAGAGAGACCAGAAGATGAAGCAACACTT 60   |  |
| DB                        | <br>403 TTAGGTGAGATGCAGAAAATTGGCTTAAGAGAGAGGCCAAGAGATGAAGCAACACTT 344   |  |
| OY                        | 61 AAGCCTTCACACCTACCTCTAAACAGTGCTTTCAAACCTCCAGTGCACAACCTGAAGCTC 120   |  |
| DB                        | 343 AAGCCTTCACACCTACCTCTAAACAGTGCTTTCAAACCTCCAGTGCACAACCTGAAGCTC 284  |  |
| OY                        | 121 TTGAAGACACTGAATATTCACACAGCAGTAGCAGTAGATGATGATGACCTTAAGGTCATT 180  |  |
| DB                        | 283 TTGAAGACACTGAATATTCACACAGCAGTAGCAGTAGATGATGATGACCTTAAGGTCATT 224  |  |
| OY                        | 181 ACCACAGGCCAGGGCGCTGGGACGCGTACTCATCATCAACCTTAATAAACGACAGCTTTGCT 240  |  |
| DB                        | 223 ACCACAGGCCAGGGCGCTGGGACGCGTACTCATCATCAACCTTAATAAACGACAGCTTTGCT 164  |  |
| OY                        | 241 TCCTCTCTAAATATAGTATCACTCAATTTTATGACACCGCAATGTTAGTACTTACTATA 300   |  |
| DB                        | 163 TCCTCTCTAAATATAGTATCACTCAATTTTATGACACCGCAATGTTAGTACTTACTATA 104   |  |
| OY                        | 301 TGCGGCTACAAAAAGSTAAACCTTTTATATTTTATATCACTTACCTTCAGCCAGCTATTGA 360   |  |
| DB                        | 103 TGCGGCTACAAAAAGSTAAACCTTTTATATTTTATATCACTTACCTTCAGCCAGCTATTGA 44  |  |
| OY                        | 361 TTATAATTAAAACTTTTCACACATATTCATTAAGTTAA 397  |  |
| DB                        | 43 TTATAATTAAAACTTTTCACACATATTAATAAAAAAAAA 7  |  |
| RESULT 2                  |   |  |
| LOCUS                     | BB223728 655 bp mRNA linear EST 31-AUG-2001   |  |
| DEFINITION                | BB223728 RIKEN full-length enriched, adult male aorta and vein Mus musculus cDNA A530083H05.3, mRNA sequence.   |  |
| ACCESSION                 | BB223728  |  |
| VERSION                   | BB223728.2 GI:15410154  |  |
| KEYWORDS                  | EST.  |  |
| SOURCE                    | house mouse.  |  |
| ORGANISM                  | Mus musculus  |  |
| REFERENCE                 | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 655)   |  |
| AUTHORS                   | Arakawa,T., Carlinucci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Koude ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasakita,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y. |  |
| TITLE                     | RIKEN Mouse ESTs (Arakawa,T. et al. 2001)   |  |
| JOURNAL                   | unpublished (2001)  |  |
| COMMENT                   | On Jul 1, 2000 this sequence version replaced gi:8892340.<br>Contact: Yoshihide Hayashizaki<br>Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute   |  |

The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenhiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-resesgsc.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subcloning of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. *Genome Res.* 10 (10), 1617-1630 (2000)  
Wagii, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Watahiki, M., Itoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura  
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer. *Genome Res.*  
10 (11), 1757-1771 (2000)  
Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara,  
Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
non-redundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)  
Yamanaka, T., Kiyosawa, H., Kondo, S., Saito, T., Shingawa, A., Atzawa,  
K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., Arakawa, T.,  
Ishii, Y. and Hayashizaki, Y.  
Mapping of 19032 mouse cDNAs on mouse chromosomes. *J. Struct.  
Funct. Genomics* 2 pre, L72-L86 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
further details.  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

[illegible]





[illegible]

|                       |  |  |     |  |  |  |  |  |  |  |
|-----------------------|--|--|-----|--|--|--|--|--|--|--|
| JOURNAL               | Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage<br>BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr<br>- Web : www.genoscope.cns.fr)<br>Determination of this BAC-end sequence was carried out as part of a<br>collaboration with the European Drosophila Genome Project (EDGP) -<br>http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC<br>library (Dros BAC) was made by Alain Billaud at CEPH (Centre<br>d'Etude du Polymorphisme Humain) with funding provided by a MRC<br>project grant. The DNA was prepared from embryos by Alain Bucheton<br>and Genevieve Payan. It has been constructed in the vector<br>pBelobAC11.  |  |     |  |  |  |  |  |  |  |
| FEATURES              | Location/Qualifiers  |  |     |  |  |  |  |  |  |  |
| source                | 1. 1101<br>/organism="Drosophila melanogaster"<br>/db_xref="taxon:7227"<br>/clone="BACN16D22"<br>/clone_lib="DrosBAC"<br>/plasmid="pBelobAC11"<br>/note="end : 17"   |  |     |  |  |  |  |  |  |  |
| BASE COUNT            | 203 a 220 c 84 g 158 t 436 others  |  |     |  |  |  |  |  |  |  |
| ORIGIN                |  |  |     |  |  |  |  |  |  |  |
| Query Match           | 10.4% Score 46.4; DB 17; Length 1101;  |  |     |  |  |  |  |  |  |  |
| Best Local Similarity | 32.2% Pred. No. 0.33; Mismatches 82; Indels 0; Gaps 0;   |  |     |  |  |  |  |  |  |  |
| Matches               | 65; Conservative 55; Mismatches 82; Indels 0; Gaps 0;  |  |     |  |  |  |  |  |  |  |
| OY                    | 243  | TCCTCTAAATGAGTTCACATCTTAAATGACACCTGATGATGATGATGATG | 302 |  |  |  |  |  |  |  |
| Db                    | 1002   | TTTATATGAAATGTTATGATGATGATGATGATGATGATGATGATGATG   | 943 |  |  |  |  |  |  |  |
| OY                    | 303  | CCGCTACAAAAGTAAAGCTTTTATATGTTTATATGATTAATCACTTAC   | 362 |  |  |  |  |  |  |  |
| Db                    | 942  | AAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATG    | 883 |  |  |  |  |  |  |  |
| OY                    | 363  | TAAATTAACATTTTCACACATCAATCAATCAATCAATCAATCAATCA    | 422 |  |  |  |  |  |  |  |
| Db                    | 882  | WHAAMWTATYAAAMWMAATATGTTATGATTAATGATTAATGATTAAT    | 823 |  |  |  |  |  |  |  |
| OY                    | 423  | TTCTTTCCCTGCTTATGAAAA                              | 444 |  |  |  |  |  |  |  |
| Db                    | 822  | TWAAAAAATAATTAATAAATA                              | 801 |  |  |  |  |  |  |  |
| RESULT 8              |  |  |     |  |  |  |  |  |  |  |
| CNS002JY              |  |  |     |  |  |  |  |  |  |  |
| LOCUS                 | CNS002JY 914 bp DNA linear GSS 26-JUL-1999   |  |     |  |  |  |  |  |  |  |
| DEFINITION            | Drosophila melanogaster genome survey sequence Sp6 end of BAC<br>BACN02L18 of DrosBAC library from Drosophila melanogaster (fruit<br>fly) genomic survey sequence.   |  |     |  |  |  |  |  |  |  |
| ACCESSION             | AL097768   |  |     |  |  |  |  |  |  |  |
| VERSION               | AL097768.1 GI:5609379  |  |     |  |  |  |  |  |  |  |
| KEYWORDS              | GSS.   |  |     |  |  |  |  |  |  |  |
| SOURCE                | Drosophila melanogaster.   |  |     |  |  |  |  |  |  |  |
| ORGANISM              | Drosophila melanogaster.<br>Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;<br>Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;<br>Ephydroidea; Drosophilidae; Drosophila.   |  |     |  |  |  |  |  |  |  |
| REFERENCE             | 1 (bases 1 to 914)<br>Genoscope.<br>Direct Submission<br>Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :<br>BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr<br>- Web : www.genoscope.cns.fr)<br>Determination of this BAC-end sequence was carried out as part of a<br>collaboration with the European Drosophila Genome Project (EDGP) -<br>http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC<br>library (Dros BAC) was made by Alain Billaud at CEPH (Centre<br>d'Etude du Polymorphisme Humain) with funding provided by a MRC<br>project grant. The DNA was prepared from embryos by Alain Bucheton<br>and Genevieve Payan. It has been constructed in the vector<br>pBelobAC11. |  |     |  |  |  |  |  |  |  |
| AUTHORS               |  |  |     |  |  |  |  |  |  |  |
| TITLE                 |  |  |     |  |  |  |  |  |  |  |
| JOURNAL               |  |  |     |  |  |  |  |  |  |  |
| COMMENT               | Location/Qualifiers<br>1. 914  |  |     |  |  |  |  |  |  |  |
| FEATURES              | Location/Qualifiers<br>1. 914  |  |     |  |  |  |  |  |  |  |
| SOURCE                |  |  |     |  |  |  |  |  |  |  |

| BASE COUNT            | 282 a   | 51 c            | 36 g      | 138 t       | 407 others |
|-----------------------|---|-----------------|-----------|-------------|------------|
| ORIGIN                |   |                 |           |             |            |
| Query Match           | 10.3%   | Score 46.2;     | DB 17;    | Length 914; |            |
| Best Local Similarity | 26.6%   | Pred. No. 0.36; |           |             |            |
| Matches 91;           | Conservative 53;  | Mismatches 198; | Indels 0; | Gaps 0;     |            |
| 103                   | AGTGCACACGACGACCTCTTGACGACACTGGAATTTACACACAGCAGTAGCAGTGAATGC        | 162             |           |             |            |
| Db                    | 443 ANNNNAANAANNNNAANNNNAANNNNAANNNNAANNNNAANNNNAANNNNAANNNNNN      | 502             |           |             |            |
| QY                    | 163 ATGTACCTTAAGGTCATTATCCACAGCGCCAGGGCGTGGCAGCGTCTCATCAACCT        | 222             |           |             |            |
| Db                    | 503 AANNNAANAANNAANAANAANAANAANAANAANAANAANAANAANAANAANAANNNNAN     | 562             |           |             |            |
| QY                    | 223 AAAAAAGCAGAGCTTGGCTCTCTCTCTAAATAGATCTACATTTTAATGCAACCTGAA       | 282             |           |             |            |
| Db                    | 563 ANANNAANAANNN       | 622             |           |             |            |
| QY                    | 283 TGTGTAGATGTTACTATATGCGCGTACAAANAAGTAAACCTTTTATATTTTATATCA       | 342             |           |             |            |
| Db                    | 623 TGMHTATTTTCTTTTAAAT       | 682             |           |             |            |
| QY                    | 343 ACTTCAGCAGCAGTGTAT        | 402             |           |             |            |
| Db                    | 663 YHAT        | 742             |           |             |            |
| QY                    | 403 TTATTTTCTAATGTGCTAGTCTTCTCCCTGCTTAATGAAA                        | 444             |           |             |            |
| Db                    | 743 TWAATTTTWTAAAT          | 784             |           |             |            |
| RESULT 9              |   |                 |           |             |            |
| LOCUS                 | CNS01720  |                 |           |             |            |
| DEFINITION            | CNS01720 1101 bp DNA linear GSS 26-JUL-1999                         |                 |           |             |            |
| REFERENCE             | Drosophila melanogaster genome survey sequence Sp6 end of BAC       |                 |           |             |            |
| AUTHORS               | BACN37P08 of DrosBAC library from Drosophila melanogaster (fruit    |                 |           |             |            |
| TITLE                 | fly), genomic survey sequence.                                      |                 |           |             |            |
| JOURNAL               | AL108704  |                 |           |             |            |
| COMMENT               | Accession   |                 |           |             |            |
|                       | AL108704.1 GI:5629008   |                 |           |             |            |
|                       | GSS.  |                 |           |             |            |
|                       | Drosophila melanogaster.  |                 |           |             |            |
|                       | Drosophila melanogaster.  |                 |           |             |            |
|                       | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;       |                 |           |             |            |
|                       | Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;          |                 |           |             |            |
|                       | Ephydroidea; Drosophilidae; Drosophila.                             |                 |           |             |            |
|                       | 1 (bases 1 to 1101)   |                 |           |             |            |
|                       | Genoscope.  |                 |           |             |            |
|                       | Direct Submission   |                 |           |             |            |
|                       | Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : |                 |           |             |            |
|                       | BP 191 J1006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  |                 |           |             |            |
|                       | - Web : www.genoscope.cns.fr)                                       |                 |           |             |            |
|                       | Determination of this BAC-end sequence was carried out as part of a |                 |           |             |            |
|                       | collaboration with the European Drosophila Genome Project (EDGP) -  |                 |           |             |            |
|                       | http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC       |                 |           |             |            |
|                       | library (Dros BAC) was made by Alain Billand at CEPH (Centre        |                 |           |             |            |
|                       | d'Etude du Polymorphisme Humain) with funding provided by a MRC     |                 |           |             |            |
|                       | project grant. The DNA was prepared from embryos by Alain Bucheton  |                 |           |             |            |
|                       | and Genevieve Payan. It has been constructed in the vector          |                 |           |             |            |
|                       | pBelobAC11.   |                 |           |             |            |
| FEATURES              |   |                 |           |             |            |
| Source                | Location/Qualifiers   |                 |           |             |            |
|                       | 1..1101   |                 |           |             |            |
|                       | /organism="Drosophila melanogaster"                                 |                 |           |             |            |
|                       | /db_xref="taxon:7227"   |                 |           |             |            |
|                       | /clone="BACN37P08"  |                 |           |             |            |
|                       | /clone_lib="DrosBAC"  |                 |           |             |            |
|                       | /plasmid="pBelobAC11"   |                 |           |             |            |
|                       | /plasmid="pBelobAC11"   |                 |           |             |            |

[illegible]

Db 697 AMTTTWTGTTTTTTTTTMMATDAMCAAMCMHACTTAACAYCAWCAHAMMTTAMNMCMCH 756

Qy 314 AGCTAAACCTTTTATATTTTATATCACTCAGCCAGCTTGATATTAATTAACA 373

Db 757 ACCTTMAATMTTMAATMTTMAATCACTTTTATTTTATTCMTATMTTDAATTAACA 816

Qy 374 TTTTCACAAATACAAATTAAGTTAACTATTTTATTTTCTAATGTCCTAGTCTTTC 429

Db 817 TATTTATTTTATATATATATTTTAAAKAKASAMATMTTMTTMTTMTTWTW 872

RESULT 11  
LOCUS BG603853 583 bp mRNA linear EST 14-AUG-2001  
DEFINITION EST052943 Plasmodium yoelii sporozoite cDNA Plasmodium yoelii cDNA  
ACCESSION BG603853  
VERSION BG603853.1 GI:15153867  
KEYWORDS EST.  
SOURCE Plasmodium yoelii.  
ORGANISM Plasmodium yoelii.  
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
AUTHORS Kappe,S.H.I., Gardner,M.J., Brown,S.M., Ross,J., Matuschewski,K., Ribeiro,J.M., Adams,J.H., Quackebush,J., Cho,J., Carucci,D.J., Hoffman,S.L. and Nussenzweig,V.  
TITLE Exploring the transcriptome of the malaria sporozoite stage  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9895-9900 (2001)  
MEDLINE 21396555  
COMMENT Contact: Malcolm J. Gardner  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301 838 3519  
Fax: 301 838 0208  
Email: gardner@tigr.org  
Request for clones, please contact: Stefan Kappe,  
kappes01@popmail.med.nyu.edu Michael Heidelberger Division,  
Department of Pathology New York University School of Medicine.

FEATURES  
source location/Qualifiers  
1..583  
/organism="Plasmodium yoelii"  
/strain="17XNL"  
/db\_xref="taxon:5861"  
/clone="PYCDV83"  
/clone\_lib="Plasmodium yoelii sporozoite cDNA"  
/dev\_stage="sporozoites from salivary gland"  
/lab\_host="E. coli TOP10"  
/note="Vector: PCR4; TA cloning; Plasmodium yoelii sporozoite cDNA library from salivary gland sporozoites 14 days post-infection"

BASE COUNT 220 a 73 c 59 g 231 t  
ORIGIN

Query Match 10.1%; Score 45.2; DB 12; Length 583;  
Best Local Similarity 52.1%; Pred. No. 0.6; Mismatches 93; Indels 0; Gaps 0;  
Matches 101; Conservative 0;

Db 467 ATGCAGACATTTT 480

RESULT 12  
LOCUS A2541515 874 bp DNA linear GSS 14-NOV-2000  
DEFINITION ENF0607F Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.  
ACCESSION A2541515  
VERSION A2541515.1 GI:11149553  
KEYWORDS GSS.  
SOURCE Entamoeba histolytica.  
ORGANISM Entamoeba histolytica.  
REFERENCE Eukaryota; Entamoebidae; Entamoeba.  
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.  
TITLE Determination of clone end sequences from Entamoeba histolytica HMI:IMSS sheared DNA library  
JOURNAL Unpublished (2000)  
COMMENT Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjl@loftus@tigr.org  
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library  
Seq primer: M13-Forward  
Class: Shotgun

FEATURES  
source location/Qualifiers  
1..874  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"  
/note="Vector: pGEM1; site1; Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 275 a 84 c 89 g 426 t  
ORIGIN

Query Match 10.0%; Score 44.6; DB 17; Length 874;  
Best Local Similarity 51.2%; Pred. No. 0.91; Mismatches 99; Indels 0; Gaps 0;  
Matches 104; Conservative 0;





DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

# FEATURES

source

Location/Qualifiers  
1..829

/organism="Brassica oleracea"

/strain="TO100DH3"

/db\_xref="taxon:3712"

/clone="BOHJ748"

/clone\_lib="BOHJ"

/note="Vector: PHOS1, Site\_1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 315 a 86 c 102 g 326 t

ORIGIN

Query Match 9.9%; Score 44.2; DB 17; Length 829;

Best Local Similarity 58.9%; Pred. No. 1.1;

Matches 76; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 281 AATGTTAGATAGTTACTATATGCGCGCTACAAAAGTAAAGTAAACCTTTTATATTTTATACAT 340

Db 751 AATTTTAGAAGACTATAATATGTACATCCCAATTAATATATATATATATCTTATTA 692

QY 341 TAACCTCAGCCAGCTATTTGATATAATTAACAATTTTCACACAATACATAGTTAACTA 400

Db 691 TATTTTCGCTAGCTATTTGAATTAAGAAATACAAATTTAATATATATAAATAATTAACA 632

QY 401 TTTTATTTT 409

Db 631 TAATAAATTT 623

Search completed: February 24, 2003, 16:31:09  
Job time : 669.074 secs

